687 2017-09-21

Note Title 9/21/2017

First, a quiz.

Where we are:

1) Describing environments
2) Black Box Optimization (BBO)
   ↳ Does not leverage MDP structure
3) Value functions - A tool for leveraging MDP structure, but not an agent on its own.

State-Value Function

$v^\pi : \mathcal{S} \to \mathbb{R}$

$$v^\pi(s) \triangleq E\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k} \,\middle|\, S_t = s, \pi\right]$$

so can just set $t = 0$ + will get same thing

Expected discounted return from state s under policy π.

- Depends on π
- Does not depend on t

$$J(\pi) = E\left[\sum_{t=0}^{\infty} \gamma^t R_t \,\middle|\, \pi\right]$$

$$= \sum_{s \in \mathcal{S}} d_0(s) \cdot E\left[\sum_{t=0}^{\infty} \gamma^t R_t \,\middle|\, \pi, S_0 = s\right]$$

$$= \sum_{s \in \mathcal{S}} d_0(s) \cdot v^\pi(s)$$

"value of state s"

$$v^\pi(s) = \sum_{k=0}^{\infty} \gamma^k E\left[R_t \,\middle|\, S_t = s, \pi\right]$$

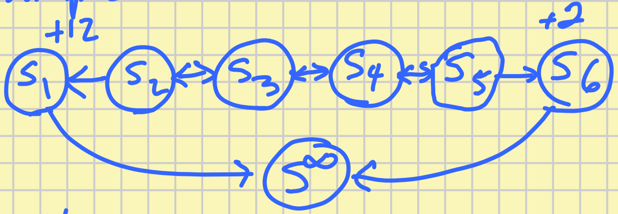
$$= \gamma^0 \sum_a \pi(s,a) \sum_{s'} P(s,a,s') R(s,a,s')$$

$$+ \gamma^1 \sum_a \pi(s,a) \sum_{s'} P(s,a,s') \sum_{a'} \pi(s',a') \cdot \sum_{s''} P(s',a',s'') R(s',a',s'')$$

$$+ \gamma^2 \cdots$$

t does not show up anywhere here

Example:



$\gamma = 0.5$

π_1 : left always

π_2 : right always

$v^{\pi_1}(s_1) = 0$

$v^{\pi_1}(s_2) = 12\gamma^0 + 0\gamma^1 + 0 \ldots\ldots = 12$

$v^{\pi_1}(s_3) = 0\gamma^0 + 12\cdot\gamma^1 + 0 \ldots\ldots = 6$

$v^{\pi_1}(s_4) = 3$

$v^{\pi_1}(s_5) = 1.5$

$v^{\pi_1}(s_6) = 0$

$v^{\pi_2}(s_1) = 0$

$v^{\pi_2}(s_2) = 0\gamma^0 + 0\gamma^1 + 0\gamma^2 + 2\gamma^3 + 0 \ldots = \frac{1}{4}$

$v^{\pi_2}(s_3) = \frac{1}{2}$

$v^{\pi_2}(s_4) = 1$

$v^{\pi_2}(s_5) = 2$

$v^{\pi_2}(s_6) = 0$

Action-Value Functions (a.k.a. State-Action-Value Functions, or Q-functions)

$q^\pi : \mathcal{S} \times \mathcal{A} \Rightarrow \mathbb{R}$ ← Again, depends on π but <u>not</u> on t.

$$q^\pi(s,a) \triangleq E\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k} \,\middle|\, S_t = s, A_t = a, \pi\right]$$

Expected discounted return of taking action a in state s and thereafter following π.

Note: $v^\pi(s) = \sum_a \pi(s,a)\, q^\pi(s,a)$

$q^{\pi_1}(s_1, L) = 0$	$q^{\pi_1}(s_1, R) = 0$
$q^{\pi_1}(s_2, L) = 12$	$q^{\pi_1}(s_2, R) = 3$ ← $\gamma \cdot 6$
$q^{\pi_1}(s_3, L) = 6$	$q^{\pi_1}(s_3, R) = 1.5$
$q^{\pi_1}(s_4, L) = 3$	$q^{\pi_1}(s_4, R) = 0.75$
$q^{\pi_1}(s_5, L) = 1.5$	$q^{\pi_1}(s_5, R) = 0$
$q^{\pi_1}(s_6, L) = 0$	$q^{\pi_1}(s_6, R) = 0$

The Bellman Equation for V^{π}

A self-consistency equation:

$$v^{\pi}(s) \triangleq E\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k} \,\middle|\, S_t = s, \pi\right]$$

$$= E\left[\gamma^0 R_t + \sum_{k=1}^{\infty} \gamma^k R_{t+k} \,\middle|\, S_t = s, \pi\right]$$

$$= E\left[R_t + \sum_{k=0}^{\infty} \gamma^{k+1} R_{t+k+1} \,\middle|\, S_t = s, \pi\right]$$

$$= E\left[R_t + \gamma \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,\middle|\, S_t = s, \pi\right]$$

$$= \sum_a \pi(s,a) \sum_{s'} P(s,a,s')\left[R(s,a,s') + \gamma E\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,\middle|\, S_{t+1} = s', \pi\right]\right]$$

Bellman Equation

$$= \sum_a \pi(s,a) \sum_{s'} P(s,a,s')\left[R(s,a,s') + \gamma v^{\pi}(s')\right]$$

$$= E\left[R_t + \gamma v^{\pi}(S_{t+1}) \,\middle|\, S_t = s, \pi\right]$$